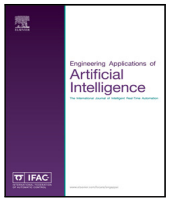
Engineering Applications of Artificial Intelligence 107 (2022) 104547

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



An efficient unsupervised image quality metric with application for condition recognition in kiln

Leyuan Wu [a], Xiaogang Zhang [a,*], Hua Chen [b], Yicong Zhou [c], Lianhong Wang [a], Dingxiang Wang [a]

[a] College of Electrical and Information Engineering, Hunan University, China
[b] College of Computer Science and Electronic Engineering, Hunan University, China
[c] The Faculty of Science and Technology, University of Macau, Taipa, Macau

ARTICLE INFO

Keywords:
Blind image quality assessment (BIQA)
Textural intensity
IQA-based application
Sintering condition recognition

ABSTRACT

In this paper, we propose an unsupervised textural-intensity-based natural image quality evaluator (TI-NIQE) by modelling the texture, structure and naturalness of an image. In detail, an effective quality-aware feature named as textural intensity (TI) is proposed in this paper to detect image texture. The image structure is captured by the distribution of gradients and basis images. The naturalness is characterized through the distributions of the locally mean subtracted and contrast normalized (MSCN) coefficients and the products of pairs of the adjacent MSCN coefficients. Furthermore, a new application pattern of image quality assessment (IQA) measures is proposed by taking the quality scores as the essential input of the recognition model. Using statistics of video quality scores computed by TI-NIQE as input features, an automatic IQA-based visual recognition model is proposed for the condition recognition in rotary kiln. Extensive experiments on benchmark datasets demonstrate that TI-NIQE shows better performance both in accuracy and computational complexity than other state-of-the-art unsupervised IQA methods, and experimental results on real-world data show that the recognition model has high prediction accuracy for condition recognition in rotary kiln.

1. Introduction

With the increasing application of machine vision technology in automatic detection and monitoring in the industrial and commercial fields, a simple and accurate image quality assessment (IQA) algorithm is critical since it can not only help for to monitor the performance of systems, but also be used as feedback to optimize vision analysis systems.

IQA can be classified into subjective and objective (Chan and Engelke, 2015). According to the information accessibility of the reference image, the objective IQA can be further divided into full-reference (FR), reduced-reference (RR) and no-reference (NR). FR and RR IQAs require reference images, which limits their application because reference images are often unavailable in practical applications (Wang, 2004; Wu et al., 2019). NR IQA, also called blind IQA (BIQA), does not need reference images in implementation. Therefore, it is more popular than FR and RR in applications. According to the distortion type of images that BIQA deals with, BIQA can be further categorized into distortion-specific and general-purpose IQAs. A distortion-specific algorithm is usually designed for one or more specific types of distortion. Through designing some target features sensitive to specific distortions, the degradation of the image is precisely quantified (Tang et al., 2015; Fang et al., 2015). In most of application scenarios, the distortion types are usually unavailable. Thus, the application scope of distortion-specific is limited. General-purpose algorithms are more applicable because they can evaluate the image quality without restriction of the distortion types. According to the accessibility of subjective scores of images, there are two strategies of general-purpose BIQA methods: supervised BIQA and unsupervised BIQA. Supervised BIQA trains a quality prediction model using large amounts of distorted images with subjective scores, and the quality score of a test image is predicted by the trained model (Mittal et al., 2012; Ma et al., 2018). The training and calibration of prediction models require plenty of image samples and subjective scores, which are time-consuming and costly acquisition tasks and usually inaccessible in many real application scenarios.

Unsupervised BIQA aims to build a reference by extracting some quality-aware features and fitting the features to a multivariate Gaussian (MVG) model with a set of pristine images. The quality of a distorted image is defined as the distance between its MVG model and the pristine MVG model (Wu et al., 2015; Zhang et al., 2015; Liu et al., 2019; Wu et al., 2020). Xue et al. proposed the quality-aware clustering (QAC) method (Xue et al., 2013) by labelling the distorted images using a FR method. In the natural image quality evaluator (NIQE) method (Wu et al., 2015) proposed by Mittal et al. the distribution parameters that are fit to the mean subtracted and contrast

* Corresponding author.
E-mail address: zhangxg@hnu.edu.cn (X. Zhang).

https://doi.org/10.1016/j.engappai.2021.104547
Received 21 December 2020; Received in revised form 9 September 2021; Accepted 4 November 2021
Available online 24 November 2021

normalized (MSCN) coefficients and the products of pairs of adjacent MSCN coefficients are employed as the quality-aware features. To boost the performance of the model, the integrated local NIQE (IL-NIQE) method (Zhang et al., 2015) proposed by Zhang et al. incorporated additional statistical features, including gradient magnitude, response of log-Gabor filters and colour, into MVG modelling. Liu et al. proposed the structure, naturalness and perception quality NIQE (SNP-NIQE) method (Liu et al., 2019) by introducing high-level natural scenes statistics (NSS) features. Wu et al. proposed the quaternion NIQE (Q-NIQE) method (Wu et al., 2020) by representing the image as a quaternion and extracted some quality-aware features from it. They also proposed a visual perception NIQE method by introducing the global perception into IQA modelling (Wu et al., 2021).

Free from the requirement of subjective scores for training, the unsupervised model affords more applicability. However, there are still some issues to be solved in the application of the unsupervised methods. One of the urgent problems to be solved is that although the prediction accuracy of unsupervised method is boosted with complex and advanced features, the computational complexity also increases significantly. Developing a both fast and effective unsupervised BIQA method for real-time applications is urgently needed. In this paper, focusing on the efficiency and effectiveness requirements of real applications, we developed a fast and simple general-purpose unsupervised IQA metric. Specifically, in addition to extracting the traditional MSCN and gradient statistical features for structure distortion from the spatial domain, we proposed an effective feature named textural intensity (TI) in the singular value decomposition (SVD) transform domain to capture the texture distortion in the image. Different from existing methods that adopt the Gabor filter to perceive the texture from various directions and scales (Zhang et al., 2015; Wu et al., 2020), this method extracts the texture from the patches, which is more efficient than existing methods. The proposed TI-NIQE is evaluated on the benchmark databases. The validation results show that the proposed metric has higher prediction accuracy and lower computational complexity than state-of-the-art (SOTA) unsupervised methods, which are more conducive to practical applications.

In previous IQA-based applications, IQA measures always play a supporting and auxiliary role in application systems. The systems select or optimize the core algorithm modules according to the feedback signals the IQA provides. The IQA-based applications are employed in two modes (Wang, 2011). In the first mode, the IQA measure acts as a selector. Besides helping select high quality images to enhance the quality-of-experience of users in multimedia systems (Fong et al., 2019), IQA helps determine the best parameters automatically for image/video processing algorithms, such as the parameter determination in image enhancement (Gu et al., 2016). The IQA can even be used to select the best image processing algorithm that generates the best perceptual quality in visual-based application systems (Wang, 2011). In the second mode, the quality score acts as a controller to control and optimize the operation of the system. For example, the IQA measure is used in an iterative mode to create feedback signals that help to update the image processing module (Yousaf and Qin, 2015), or is implemented in the core of the optimization algorithm in the image processing algorithm (Preiss et al., 2014).

In this paper, we apply the IQA measure to an application in a completely different way than previously implemented by taking the evaluation quality score as the essential input of the pattern recognition system. The application scenery is a vision-based condition recognition system of rotary kilns. The video quality is affected by coal and smoke/dust in industrial sites, while the smoke/dust concentration and the flicker degree of the video image are different under different conditions. By analysing the statistical characteristics of the image quality scores of flame videos, an IQA-based condition recognition system of rotary kilns is proposed in this paper.

Our contributions of this paper can be summarized as follows. First, we propose an effective TI-based NIQE (TI-NIQE) method by texture, structure and naturalness measurements. It is more effective than the SOTA unsupervised methods with higher prediction accuracy and lower computational complexity on benchmark databases. Second, considering that the rotary kiln has different video qualities and flickers under different working conditions, a condition recognition model using statistics of video quality scores by TI-NIQE is designed. The validation results on real-word data demonstrate that the model outperforms other methods. Finally, taking the quality scores as the essential input of the recognition model, we propose a new application pattern of IQA measures for vision-based recognition systems, which provides an instructive example for the IQA-based applications.

The rest of the paper is organized as follows. In next section, we present the proposed unsupervised BIQA model. In Section 3, we describe the condition recognition model using the quality score sequence. In Section 4, we show the experimental results on benchmark datasets and real word data, and we conclude this paper in Section 5.

2. The proposed unsupervised BIQA method

A high-quality image possesses certain regular statistical properties, while the degradation will break the regularity. Therefore, the statistics of a distorted image will be measurably different from those of pristine images. The BIQA methods include two procedures: perceived model construction and distance computation. The construction of the pristine model and distortion model is the same: the images are first divided into patches, and a visual-feature vector is extracted from each patch. Then, the feature vectors are stacked together and fitted with a MVG model.

Quality-aware feature extraction is the key issue in the BIQA algorithm, and the selection of features should be an effective representation of visual quality variations. The human visual system (HVS) is highly sensitive to the structure and texture information in visual scenes (Wang, 2004). Therefore, the variation in structure and texture (smoothness/roughness) can reflect the quality degradation degree well. Based on this concern, we design our BIQA features by capturing the structure and texture information in the image.

2.1. Quality-aware features extraction

A 2-D transform SVD decomposes the image into several basis images weighted by transformation coefficients, and visual quality can be assessed by the changes in basis images and transformation coefficients. For an image $I_{m\times n}$, SVD decompose it into three parts: left singular vector $U_{m\times m}$, right singular vector $V_{n\times n}$ and singular values $\delta_{k\times k}$ as follows:

$$I_{m\times n} = [u_1, .u_i., u_m] \begin{bmatrix} \delta_1 & & \\ & \ddots & \\ & & \delta_k \end{bmatrix} [v_1, .v_i., v_n]^T \tag{1}$$

where $k = min(m, n)$. The matrix $u_i v_i^T$ is the basis image of I, which denotes the low frequency (major structure) with small i, and denotes high frequency (finer details) with large i. The sum of the basis images forms the whole structure of the image, and is formulated as (Narwaria and Lin, 2012):

$$s = \sum_{i=1}^{k} u_i v_i^T \tag{2}$$

To visually show this point, Fig. 1(a) illustrates a clear image, and (b) and (c) are the Addictive Gaussian Noise (AGN) — distorted, Gaussian blur (GB)-distorted versions of (a), respectively. Fig. 1(d) illustrates the s map of (a). Notably that the s map can effectively capture the structure of the image. Any changes in the image will cause the changes in U and V according to perturbation theory, and distortions affect the structure of the visual perception represented by the basis image. Fig. 1(e) illustrates the s distribution of (a)~(c), and it shows that the s distribution can be accurately fitted with the zero-mean generalized

Fig. 1. (a) is a clear image, (b) and (c) are the white noise (WN) and Gaussian blur (GB) distorted of (a). (d) illustrates the s map of (a), and (e) illustrates the s distribution of (a)~(c).

Gaussian distribution (GGD) (Sharifi and Leon-Garcia, 1995) with two parameters α and β:

$$f_{GGD}(\chi;\alpha,\beta)=\frac{\alpha}{2\beta\Gamma(1/\alpha)}\exp\left(-\left(\frac{\chi}{\beta}\right)^{\alpha}\right) \tag{3}$$

where $\Gamma(\cdot)$ is the gamma function as follow:

$$\Gamma(\chi)=\int_0^{\infty}t^{\chi-1}e^{-t}\,dt,\ \chi>0 \tag{4}$$

The fitted parameters α and β are employed as the quality-aware features.

The singular value (SV) δ_i is the weight of basis image $u_i v_i^T$, and it is related to the luminance changes in images. Any distortions in the image will cause a change in the luminance or texture of the image, which represents a change in the singular values. In the M-SVD metric (Shnayderman et al., 2006), Shnayderman et al. defined the activity level of an image as the ratio between the largest and the second largest singular values. A higher activity level indicates a rougher or stronger texture of the image. In Wu et al. (2019), the author found the first singular value related to the mean luminance of the image, and the second and subsequent singular values were more sensitive to the contrast change of the image. The texture variation of an image is more sensitive to contrast change but less sensitive to illustration change, so we use the second and subsequent singular values to represent the variation of the image. Considering that the energy of high frequencies (SVs with large indices) will be increased when the image becomes rougher, and is reflected in the SVs with large indices, we define the textural intensity (TI) as the proportion of high frequency components in the image:

$$TI=\frac{\sum_{i=4}^{k}\delta_i}{\sum_{i=2}^{k}\delta_i+c_1} \tag{5}$$

where the sum from the fourth to the last SVs represents the TI of high frequency components of image, and the sum from the second to the last SVs represents the TI of whole image, and c_1 is a constant value to increase the stability of TI. A larger TI value indicates the rougher of an image. To validate this point, we calculate TI values on an reference image (Fig. 2(a)) and five level AGN distorted images (Fig. 2(b)~(f)) as well as five level GB distorted images (Fig. 2(g)~(k)). To capture the local information, the image is decomposed into non-overlapped 8×8 patches. The mean value of TI on each patch is set as the measure index. The image becomes rougher with the increase noise level, and smoother with the increase of blur level.

Fig. 3(a) illustrates the TI results of the images illustrated in Fig. 2. Note that the distortion level 0 denotes the clear image. The figure shows that the TI value increased with the increasing AGN distortion level, and decreased with the increasing GB distortion level. Therefore, the TI value reflects the textural level of the image, since the AGN-distorted image is rougher than GB-distorted. Fig. 3(b) shows the absolute difference value of TI between the clear image and the distorted image in Fig. 3(a): $D=|TI_{clear}-TI_{distorted}|$. It can be found that the D value monotonically increased with the increasing of distortion level. That indicates TI can effectively capture the distortions in the image.

In addition, motivated by the fact that HVS is sensitive to the most distorted areas, the percentile strategy is adopted to improve the correlations with subjective perception. Specifically, TIs are calculated on the non-overlapped patches across the image. The quality-aware features are calculated as the average value of the lowest 10th percentile and the 100th percentile TI values.

In regard to the naturalness of the images, modelling by the locally MSCN coefficients and the products of pairs of adjacent MSCN coefficients, has been proven to be an effective measure for naturalness measurement (Wu et al., 2015; Zhang et al., 2015; Liu et al., 2019; Wu et al., 2020). Given an image I with spatial coordinates i and j, the MSCN is calculated as:

$$M(i,j)=\frac{I(i,j)-u(i,j)}{\sigma(i,j)+c_2} \tag{6}$$

where c_2 denotes a constant value, and is fixed as 1. u and σ are the local mean and contrast, which can be obtained through

$$\mu(i,j)=\sum_{m=-M}^{M}\sum_{n=-N}^{N}w_{i,j}I(i+m,j+n) \tag{7}$$

$$\sigma(i,j)=\sqrt{\sum_{m=-M}^{M}\sum_{n=-N}^{N}w_{m,n}(I(i+m,j+n)-\mu(i,j))^2} \tag{8}$$

For a clear image, the MSCN coefficients are subjective to a GGD. The fitting parameters α and β will be changed accordingly when the image is affected by distortions (Wu et al., 2015; Zhang et al., 2015; Liu et al., 2019; Wu et al., 2020). The fitting parameters α and β are employed as quality-aware features.

The fitting parameters that correspond to adjacent MSCN coefficients products are also selected to characterize the naturalness of an image. For a MSCN map $M(i,j)$, the adjacent MSCN map at the horizontal, vertical, main-diagonal, secondary diagonal orientations can be calculated as $M(i,j)M(i+1,j)$, $M(i,j)M(i,j+1)$, $M(i,j)M(i+1,j+1)$, and $M(i,j)M(i+1,j-1)$. The adjacent MSCN coefficients products can be well modelled using a zero mode asymmetric GGD (AGGD) (Lasmar et al., 2009):

$$f_{AGGD}(\chi;\gamma,\beta_l,\beta_r)=\begin{cases}\frac{\gamma}{(\beta_l+\beta_r)\Gamma\left(\frac{1}{\gamma}\right)}\exp\left(-\left(\frac{-\chi}{\beta_l}\right)^{\gamma}\right) & \chi\le 0\\ \frac{\gamma}{(\beta_l+\beta_r)\Gamma\left(\frac{1}{\gamma}\right)}\exp\left(-\left(\frac{\chi}{\beta_r}\right)^{\gamma}\right) & \chi>0\end{cases} \tag{9}$$

with the mean of the distribution is:

$$\eta=(\beta_r-\beta_l)\frac{\Gamma\left(\frac{2}{\gamma}\right)}{\Gamma\left(\frac{1}{\gamma}\right)} \tag{10}$$

The fitting parameters $(\gamma,\beta_l,\beta_r,\eta)$ are obtained on each adjacent MSCN map. Therefore 16 features are obtained. In totally, 18 features are employed to characterize the naturalness of an image.

A high-quality image usually has high contrast and clear edges, which can be captured by the gradient map. The fitting parameters of the gradient map are employed as quality-aware features. The gradient map in horizontal G_h and in vertical G_v are obtained by convolving the image with two Gaussian derivative filters along the horizontal f_h and vertical f_v directions. For a zero-mean Gaussian distribution $g(x,y)=\frac{1}{2\pi\sigma^2}e^{-(x^2+y^2)/(2\sigma^2)}$, f_h and f_v are defined as:

$$f_h=\frac{\partial}{\partial x}g(x,y)=\frac{-x}{2\pi\sigma^4}e^{-\frac{(x^2+y^2)}{(2\sigma^2)}} \tag{11}$$

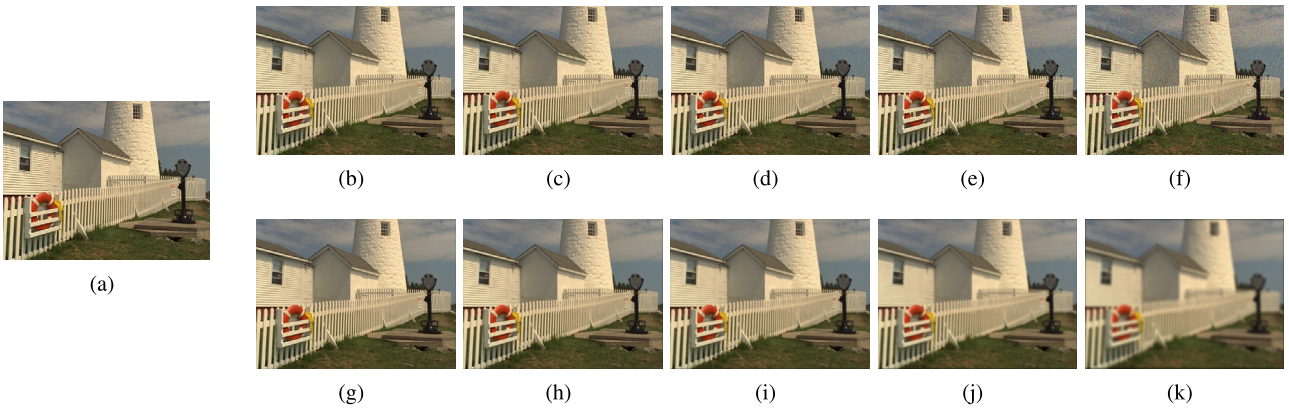


Fig. 2. (a) is a clear image from TID2013 database. (b)~(f) are AGN distorted images, and (g)~(k) are GB distorted images. The distortion level increased from left to right.

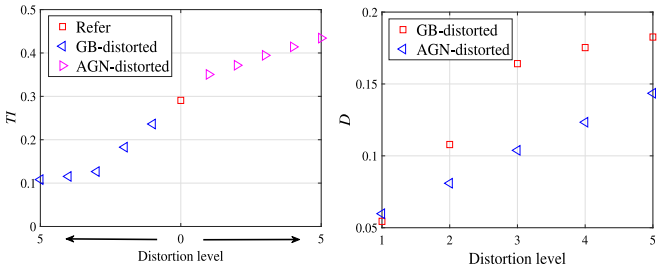


Fig. 3. (a): TI values of the images in Fig. 2. (b): D values of the GB-distorted and AGN-distorted images with different level.

$$f_v = \frac{\partial}{\partial y} g(x,y) = \frac{-y}{2\pi\sigma^4} e^{-\frac{(x^2+y^2)}{(2\sigma^2)}} \tag{12}$$

The gradient coefficients G_h and G_v can be fitted with zero mean GGD [6]. The fitted parameters α and β are employed as the quality-aware features.

In summary, 26 features are extracted from an image, with 2 features to characterize the texture, 18 features to characterize the naturalness and 6 features to characterize the structure. Furthermore, the features are extracted in the original scale and the downsampled scale, as HVS perceives the image in a multiscale strategy. Therefore, a total of 52 features are employed for IQA modelling.

2.2. Feature pooling and quality score computation

To capture the local information of an image, the images are divided into non-overlapped patches (denoted as bs). The quality-aware features are extracted on the patches, and fitted with a MVG model with two fitting parameters: the mean vector u and the covariance matrix Σ, formulated as:

$$f(\chi; u, \Sigma) = \frac{1}{(2\pi)^{l/2}|\Sigma|^{1/2}} \exp\left(-\frac{1}{2}(\chi-u)^T \Sigma^{-1}(\chi-u)\right) \tag{13}$$

where χ are the extracted features and $l = 52$ is the length of the features. The pristine MVG model is first constructed by extracting quality-aware features from 125 pristine images from the NIQE model (Wu et al., 2015). The fitting parameters u_p and Σ_p are calculated. Furthermore, only the patches containing rich texture information are selected for learning. In practice, patches with a larger contrast value than the γ% peak patch contrast are selected. Then, the same process is performed on the test images to obtain the fitting parameters u_d and Σ_d. Finally, the quality score of the test image is formulated as the Bhattacharyya distance between the MVG model fitted to pristine images and the MVG model fitted to the test image.

$$Q_t = \sqrt{(\mu_t - \mu_p)^T \left(\frac{\Sigma_t + \Sigma_p}{2}\right)^{-1} (\mu_t - \mu_p)} \tag{14}$$

Note that a higher Q indicates an image with relatively lower quality.

3. IQA-based sintering condition recognition model

In most of IQA-based applications, the image quality is used as a supplementary means for the application, such as the parameter selection algorithms. In this paper, we propose to use the evaluated quality scores as the direct factors, that is, the feature inputs of the classifier, for a visual recognition application. This strategy is based on the fact that in a vision-based condition recognition of rotary kilns, different conditions have different patterns of image quality sequences in flame videos.

3.1. Background of condition recognition in rotary kiln

Monitoring and recognition of conditions based on flame images and videos have been developed in many coal-fired industries (Wang et al., 2020a,b). There are three main conditions: normal, super-chilled, super-heated according to the temperature level in the burning zone of the rotary kiln. The normal condition represents the temperature in the burning zone within the set range, and it is the desired condition. The super-heated and super-chilled conditions denote the temperatures are lower or higher than the desired temperatures, which are abnormal conditions and should be avoided in the burning process. Researchers have developed various methods to determine the conditions of the rotary kilns. They adopted the framework in which a feature extraction process followed a classification method to determine the conditions. The extracted features are the most obvious difference for different methods. There are three types of methods according to the feature extraction process (Hua Chen, 2020). The first type segments several regions of interest areas (ROIs) (e.g., flame region, material region) from the flame image, and then extracts some statistical features on it (Li et al., 2002). However, the ROIs are difficult to segment in some flame images due to the presence of dust and smoke. Hence the second type of method extracts global and local features from flame images (Wang et al., 2017). This strategy does not need the segmentation process, and therefore generates more robust performance. With the development of deep learning networks, the last type of method extracts the features automatically based on the deep-learning methods (Qiu et al., 2019). Visual-based methods have a vital defect in that the recognition performance is affected significantly by the quality of the images and videos.

According to field experience, the condition of the rotary kiln is closely related to the quality of flame videos. When the sintering

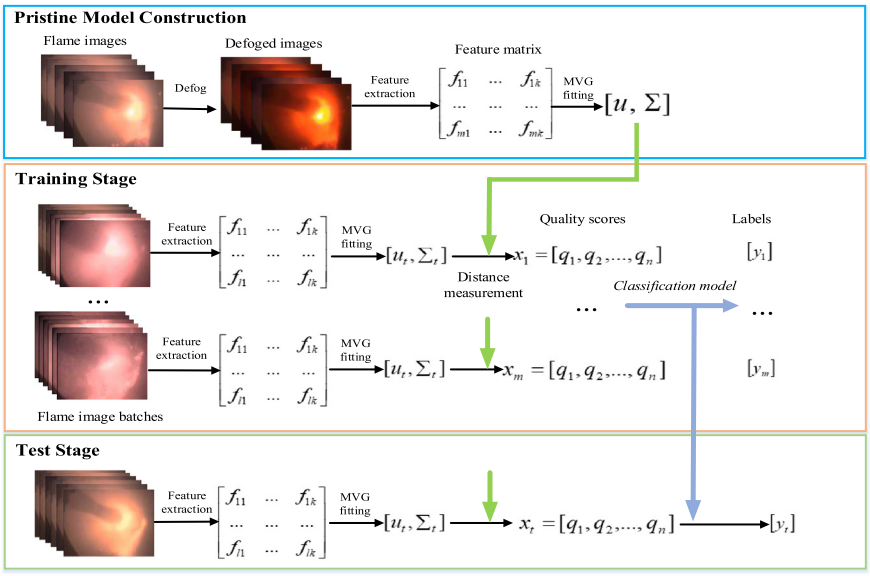


Fig. 4. The framework of the proposed quality-based condition recognition model.

temperature is moderate and the condition is normal, there is less dust and smoke in the burning zone, so images and videos captured under normal condition are of relatively high quality. Under super-chilled condition, the raw material has not sintered enough and still remains as a powder and is full of air in the burning zone, which deteriorates the quality of flame images and videos. In addition, the flame videos flicker seriously, which makes the quality of the flame image unstable. Under super-heated condition, a high temperature results in a large area of intensity saturation in the image, similar to over-exposed images (Chen et al., 2016). This makes the quality of flame images relatively poor, but the qualities of images remain relatively stable. Under different sintering conditions, the quality of the flame videos exhibits different patterns. Therefore, we attempt to use the quality of flame video as the visual feature to recognize the condition of the kiln.

3.2. Framework of the condition recognition model

The framework of the condition recognition model is illustrated in Fig. 4. First, some flame images with relatively high quality are employed to build a pristine MVG model. Specifically, the flame images are defogged to obtain the pristine flame images. The quality-aware features are extracted from defogged images and were fitted with a MVG pristine model. Then, in the training stage, the flame image sequence is first extracted from the flame video and is divided into serial batches. Each batch contains n flame images. Then, for each flame image, quality-aware features are extracted and fitted with a MVG model. The quality of an image is calculated as the Bhattacharyya distance between the MVG pristine model and the MVG model. The quality of each image is obtained and is combined into a quality vector $x = [q_1, q_2, \ldots, q_n]$. Finally, a classification model is trained to map the feature vectors to the conditions y ($y = 1, 2, 3$) of the kiln. In this paper, the multi-class optimal margin machine (mcODM) (Zhang and Zhou, 2017) is employed as the classification model. Finally, in the test stage, similar to the process of the training stage, a quality vector $x_t = [q_1, q_2, \ldots, q_n]$ is extracted from the testing flame image batch and fed into the trained classification model to obtain the classification result y_t.

3.3. Pristine model construction of flame images

We attempted to use the pristine MVG model learned from natural images to compute the image qualities of flames. However, the recognition performance was not satisfactory (shown in Table 5). We surmise that this occurred because the model was constructed using pristine natural images under the natural light source. Nevertheless, the flame image is slightly different from the natural image. The light source of the flame image in the burning zone is the burning material and flame image in the centre of the image, which is not a natural light source. Therefore, we train a pristine MVG model using a set of high-quality flame images. Because of the on-site combustion of coal and material, pristine flame images cannot be obtained, as dust and smoke are always present in the flame images. Therefore, we use the guided filter algorithm (He et al., 2013) for enhancement of flame images to obtain pristine flame images. Fig. 5 illustrates a flame image and its filtered image. The flame image is effectively deblurred, and its structure and edge are retained. We used a total of 108 filtered flame images for fitting with a pristine MVG model, and these severed as an "reference" against the test flame image. The quality of the test flame image is calculated as the Bhattacharyya distance between the reference MVG model and the MVG model from the test flame image.

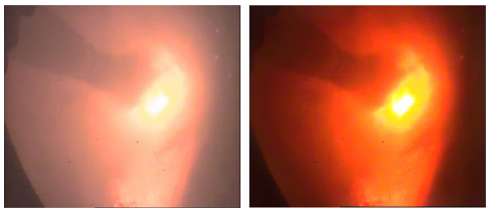
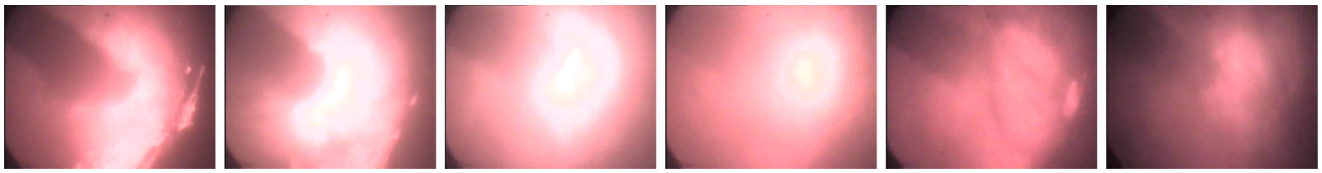


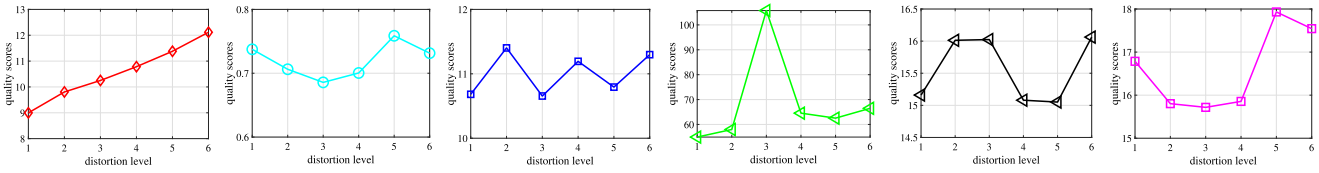
Fig. 5. (a): A flame image and (b): the corresponds filtered image.

To validate the effectiveness of our proposed IQA model on flame images, we evaluated it on six flame images with increasing distortion levels, as shown in Fig. 6. The quality scores of the images in Fig. 6 evaluated by the proposed IQA are shown in Fig. 7(a). We can see that the quality scores monotonically increase with increasing distortion level. In contrast, we compute the quality scores by existing unsupervised IQA algorithms: QAC, NIQE, IL-NIQE, SNP-NIQE and Q-NIQE. The results are plotted in Fig. 7(b)~(f). The quality scores evaluated by the competing methods do not change monotonically with the change in distortion level. This indicates that the proposed model is more effective in evaluating the distortion level of flame images compared with existing unsupervised IQA algorithms.



(a) Level1 (b) Level2 (c) Level3 (d) Level4 (e) Level5 (f) Level6

Fig. 6. Six flame images with increasing distortion level.



(a) (b) (c) (d) (e) (f)

Fig. 7. The quality scores of the flame images in Fig. 6 using the model (a) Proposed. (b) QAC, (c) NIQE, (d) IL-NIQE, (e) SNP-NIQE, (f) Q-NIQE.

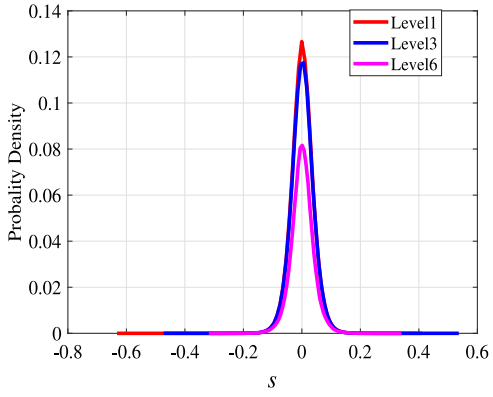


Fig. 8. The s distribution of the three flame images illustrated in Fig. 6.

Furthermore, one crucial question to be verified is whether the method used to extract feature from natural images can be used in the quality-aware feature extraction of flame images. To verify this point, Fig. 8 illustrates the s distribution of the three flame images in Fig. 6. This figure shows that the s distribution probability density can also be well modelled by the zero-mean GGD. Therefore, it is feasible to adopt feature extraction method used for natural images to extract the quality-aware features from flame images.

To investigate the relation between quality scores and conditions in rotary kiln, we perform the proposed IQA algorithm to compute quality scores on four flame clips with different sintering conditions from ZhongZhou Aluminium Corporation in China. Each video contains 5 min at 25 average frames per second fps. Therefore, each video contains $5 \times 60 \times 25 = 7500$ frames. To reduce the computational cost, we capture an frame every 6 frames for each video. Therefore, 1250 frames are analysed in each video. The results are illustrated in Fig. 9. The quality score sequences of the flame videos have different patterns under different conditions. One can observe that the image quality is high and stable under normal condition (Fig. 9(a)), therefore yielding small mean and variance values. In the super-heated condition, the image quality is relatively poor but remains stable (Fig. 9(b)), that means the mean of sequence is large while the variance is small. In the super-chilled condition, the image quality is worst and unstable (Fig. 9(c)). The mean value and the variance are both large. Specifically, in Fig. 9(d), when the condition is changed from normal to super-chilled, an increasing and unstable quality score sequence is observed. This indicates that the statistics of the quality score sequence can be used to distinguish burning conditions of rotary kilns.

4. Experimental results

4.1. Experimental protocol and setup

We comprehensively compare the proposed model with existing state-of-art unsupervised models: QAC (Xue et al., 2013), NIQE (Wu et al., 2015), IL-NIQE (Zhang et al., 2015), SNP-NIQE (Liu et al., 2019) and Q-NIQE (Wu et al., 2020) on four popular public IQA databases: LIVE (Sheikh et al., 2006), CSIQsub (Larson and Chandler, 2010), TID2013sub (Ponomarenko et al., 2013), and IVCsub (Le Callet and Autrusseau, 2005). In the comparison, the prediction accuracy and computational complexity are taken into consideration. Similar to the experiment on SNP-NIQE and Q-NIQE approaches, the most common distortions (e.g., JPEG, JP2K, AGN and GB) are selected for testing. Detailed information on the databases is shown in Table 1. Four commonly used measure indices, the Spearman Rank order Correlation coefficient (SRCC), Kendall's rank correlation coefficient (KRCC), Pearson linear correlation coefficient (PLCC) and root mean square error (RMSE), are employed for evaluation. SRCC and KRCC measure the prediction monotonicity of an IQA model. The PLCC measures the linear correlation between the objective quality scores and subjective scores. RMSE measures the errors between subjective scores and the objective scores after regression, with the regression function is modelled as follows:

$$f(\chi) = \theta_1 \left(\frac{1}{2} - \frac{1}{1 + \exp\left(\theta_2 \cdot (\chi - \theta_3)\right)} \right) + \theta_4 \cdot \chi + \theta_5 \tag{15}$$

where $\theta_{1,..5}$ are the five parameters to be fitted. Larger SRCC, KRCC and PLCC values as well as smaller RMSE value indicate better performance of the model. To obtain the overall performance of an IQA model, the direct average (denoted as avg) and weighted average (denoted as AVG) are commonly used. The number of images in the database is set as the weight.

There are three parameters in our model: the block size bs, the threshold percentile value γ and the length of the quality vector n. Fig. 10 illustrates the SRCC values on the LIVE, CSIQsub and TID2013sub databases, with bs ranging from 56 to 104 with an interval of 8, and γ ranging from 0.55 to 0.8 with an interval of 0.05. Fig. 10 indicates that a higher SRCC value can be obtained when bs is within the interval [72 96], and γ within the interval [0.55 0.7]. In our model, bs is fixed as 96, and γ is fixed as 0.6. Fig. 11 illustrates the overall prediction accuracy versus the length of the flame image sequence n, with n ranging from 10 to 50 with an interval of 10. In this experiment, we employ n flame images from the 25-fps flame video by extracting one frame in six frames. Therefore, when predicting

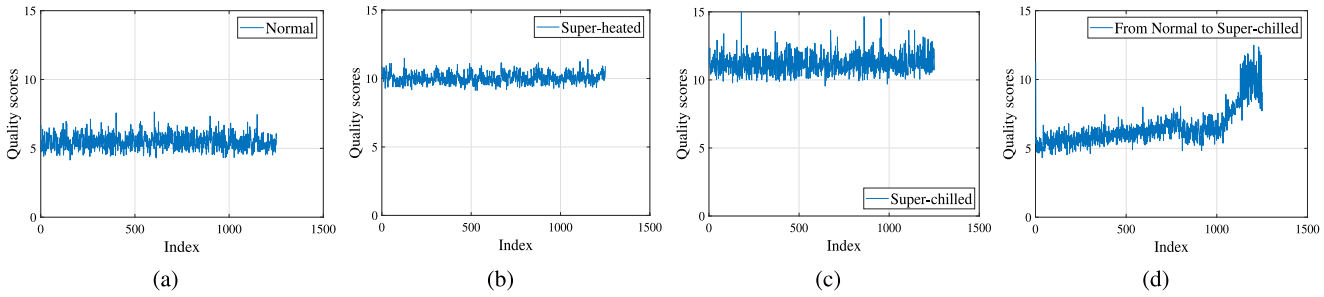


Fig. 9. The quality evaluation of the proposed model on four flame videos with the condition is (a) Normal, (b) Super-heated, (c) Super-chilled and (d) from Normal to Super-chilled.

Table 1
Main information about tested image quality databases.

Database	LIVE	CSIQsub	TID2013sub	IVCsub
Reference images Number	29	30	25	10
Distorted images Number	779	600	500	120
Distortion types	JP2K, JPEG, FF, GB, WN	JP2K, JPEG, GB, AWGN	JPEG, JP2K, AGN, GB	JP2K, JPEG, GB
Observers	161	35	917	15

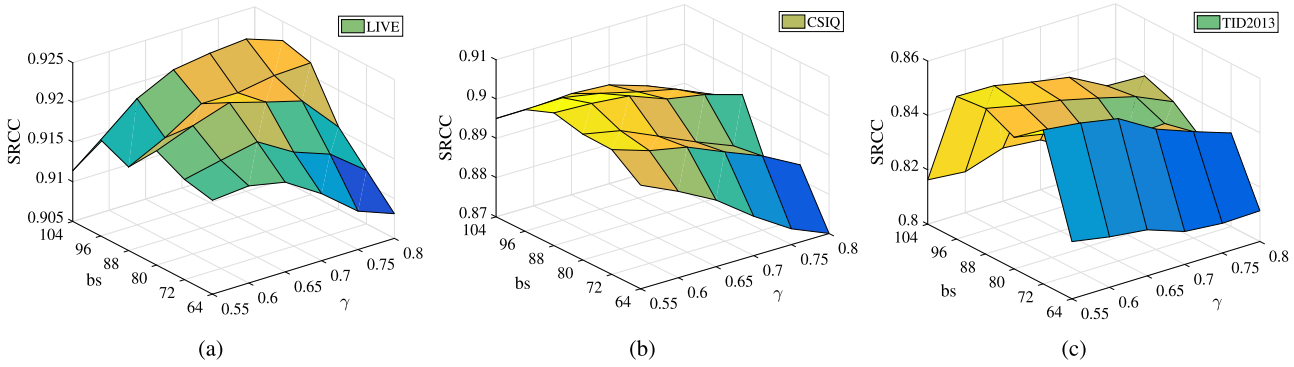


Fig. 10. The SRCC values under various parameters on the (a) LIVE database, (b) CSIQsub database, (c) TID2013sub database.

the working condition at time t, the continuous image frames of the previous $n * 6/25$ time period are used. For each experiment, the model is tested 50 times, and the average value is illustrated. When n is too small, the recognition model is susceptible to short-term abnormal working conditions (such as deflagration). When n is large enough, the model has good robustness and can effectively suppress short-term abnormal working conditions interference. One can observe that the model achieves relatively high classification accuracy when $n \geq 30$. Therefore, taking into account both calculation and accuracy, we employ $n = 30$. This means that we adopt the flame video of the previous 7 s to extract flame characteristics.

4.2. Experimental on benchmark databases

First, we analyse the extracted features. Fig. 12 shows the SRCC values on the LIVE, CSIQsub, TID2013sub and IVCsub databases when constructing the model using the structure+texture features, naturalness features, and their combination. One can observe that the prediction accuracy is boosted when combining the structural and texture features as well as the naturalness features. That indicates the features play a complementary role in image representation.

Then, the proposed method is compared with the competing methods on each database. The results are shown in Table 2. The best two results in each row are highlighted in red and blue, respectively. One can observe that the performance of the proposed model is nearly consistent well on all measure indices. Overall, the proposed model achieves the best results regardless of whether the direct average or weighted average is used.

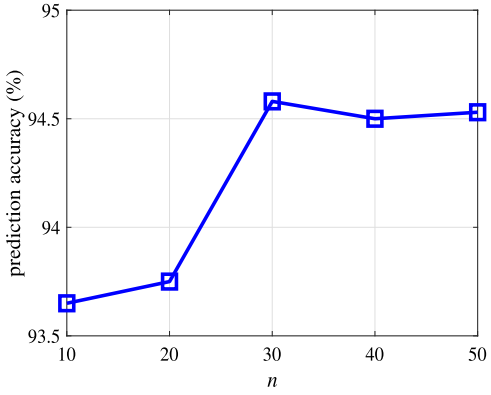


Fig. 11. The prediction accuracy versus the length of the flame image sequence n.

In practical applications, the speed of the model is also an important factor for consideration. Finally, the computational complexity of a model is evaluated in terms of the fps. This experiment is conducted on a Dell workstation with a 3.2 GHz Intel Core 7 processor and 16 GB RAM. The software platform is Matlab R2018b. Each model is tested on LIVE 20 times, and the average results are shown in Table 3. Our model is much faster than SOTA models such as IL-NIQE, SNP-NIQE and Q-NIQE. QAC and NIQE are faster than ours, but they have the modest prediction accuracy. Combining Tables 2 and 3, One can observe that the proposed model not only has higher prediction accuracy but also has lower complexity than other SOTA unsupervised methods.

Table 2
Evaluation results compared with unsupervised methods.

Database	Measure index	QAC (Xue et al., 2013)	NIQE (Wu et al., 2015)	IL-NIQE (Zhang et al., 2015)	SNP-NIQE (Liu et al., 2019)	Q-NIQE (Wu et al., 2020)	TI-NIQE
LIVE	SRCC	0.8443	0.9083	0.8978	0.9073	0.9113	0.9268
	KRCC	0.6443	0.7310	0.7129	0.7350	0.7323	0.7664
	PLCC	0.7575	0.9067	0.9025	0.9059	0.9077	0.9275
	RMSE	17.8371	11.5223	11.7686	11.5722	11.4622	10.2119
CSIQsub	SRCC	0.8364	0.8714	0.8802	0.9013	0.9046	0.8956
	KRCC	0.6272	0.6863	0.6980	0.7205	0.7216	0.7270
	PLCC	0.8474	0.8883	0.9070	0.9082	0.9126	0.9214
	RMSE	0.1500	0.1298	0.1190	0.1183	0.1157	0.1098
TID2013sub	SRCC	0.7862	0.7976	0.8420	0.8574	0.8586	0.8608
	KRCC	0.5941	0.5930	0.6536	0.6592	0.6562	0.6687
	PLCC	0.7801	0.8092	0.8582	0.8484	0.8576	0.8926
	RMSE	0.8726	0.8195	0.7161	0.7385	0.7174	0.6288
IVCsub	SRCC	0.7424	0.7911	0.8494	0.8390	0.8596	0.8923
	KRCC	0.5352	0.6029	0.6571	0.6532	0.6717	0.7111
	PLCC	0.7547	0.7971	0.8604	0.8507	0.8735	0.9066
	RMSE	0.8148	0.7499	0.6329	0.6528	0.6047	0.5241
avg	SRCC	0.8023	0.8421	0.8674	0.8763	0.8835	0.8939
	KRCC	0.6502	0.6533	0.6804	0.6920	0.6955	0.7183
	PLCC	0.7849	0.8503	0.8820	0.8783	0.8879	0.9120
AVG	SRCC	0.8213	0.8625	0.8757	0.8889	0.8930	0.8989
	KRCC	0.6980	0.6754	0.6902	0.7068	0.7064	0.7268
	PLCC	0.7900	0.8702	0.8902	0.8889	0.8946	0.9157

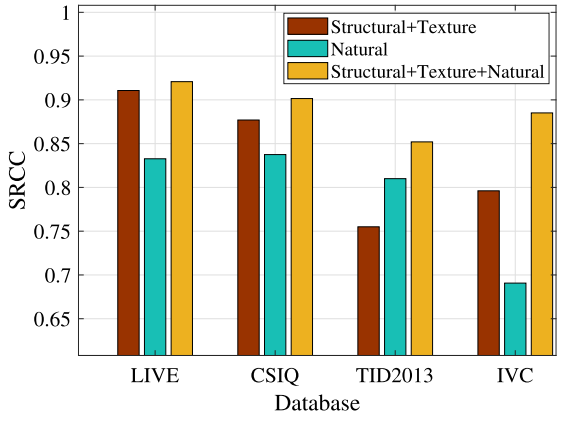


Fig. 12. SRCC values when constructing the model using the Structural and Texture features, Naturalness features and the combination of them.

Table 3
Computational complexity of the tested methods on LIVE database.

Metrics	QAC	NIQE	IL-NIQE	SNP-NIQE	Q-NIQE	TI-NIQE
fps	8.2508	9.5420	0.2729	0.5172	0.3502	2.4155

4.3. Experiments on real-word data with the condition recognition model of rotary kilns

To train and test the model, 53 2-min videos at 25-fps recorded in the ZhongZhou Aluminium Corporation in China are used, which include 20 videos in the normal condition, 20 videos in super-chilled condition and 13 videos in the super-heated condition as labelled by an experienced kilnman. There are $25 \times 2 \times 60 = 3000$ frames for each video. We capture an image every 6 frames. Therefore, 500 flame images are obtained from each video. Since we chose 30 flame images for condition recognition, we obtained 471 samples for each video. For each condition, 4 videos are used for training, and the rest are used for testing (which means a total of 12 videos for training and 41 videos for testing). Overall, 5652 samples are used for training, and 19 311 samples are used for testing. Detailed numbers of samples used for training and testing are listed in Table 4.

Table 4
The number of samples used in the sintering condition recognition.

	Normal	Super-chilled	Super-heated	Total
Training	1884	1884	1884	5652
Testing	7536	7536	4239	19311

Two types of models are employed for comparison. The first type of model comes from existing unsupervised methods (i.e., QAC, NIQE, IL-NIQE, SNP-NIQE, Q-NIQE). They are retrained using high-quality flame images and implemented in the flame image quality evaluation. The second type of model consists of two deep learning classification models: deep convolution neural networks (CNN)-based and transfer learning (TL)-based. Detailed information of the two models is illustrated as follows:

The CNN-based classification model: including an input layer, two convolution layers, two dropout layers, two ReLU layers, two cross channel normalization layers and two max pooling layers to extract the deep features. These are followed by a fully connected layer, a softmax layer and a classification layer to obtain the classification results.

TL-based classification model: a model based on the alexanet network is established. By using the transfer learning, the network model is initialized by the parameters of AlexNet's convolutional network trained on the ImageNet dataset to realize parameter migration. The model is retrained on the flame image dataset to realize the condition recognition of rotary kilns.

Table 5 illustrates the prediction accuracy of the proposed method and the competing methods. Note that only the "proposed-original" is trained on pristine natural images, and the rest are trained on high-quality flame images. Some conclusions can be drawn from Table 5. First, comparing the results of the proposed-original method and the proposed-flame method, a much better results can be achieved when the method is retrained using the flame images. Therefore, it is necessary to retrain the model using high quality flame images. Next, the proposed method achieves much better performance than existing unsupervised methods. This indicates that the proposed method is more effective in capturing the distortions in the flame image. Finally, the deep learning-based methods show good performance under normal and super-chilled conditions. However, the performance is very poor under super-heated condition. This is mainly because the flame image in the super-heated condition is not much different from that in the

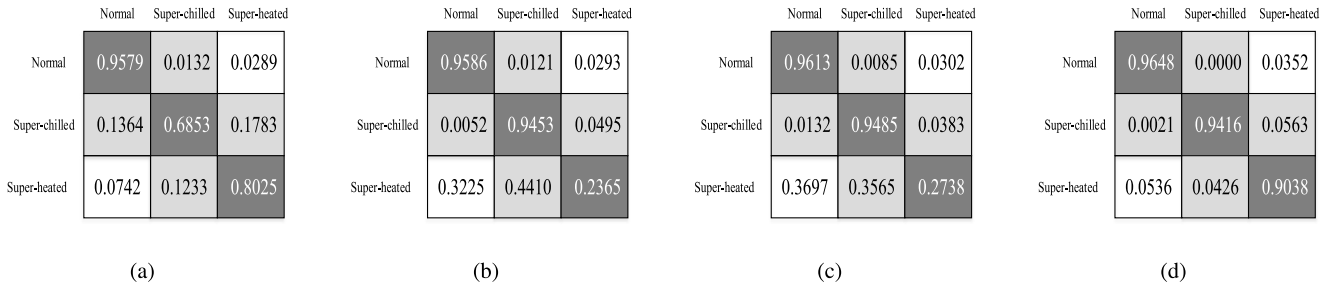


Fig. 13. The confusion matrix of the best four sintering condition recognition methods: (a) Q-NIQE, (b) CNN-based, (c) TL-based and (d) the proposed methods.

Table 5
The prediction accuracy of the proposed and the competing models.

Method	Normal	Super-chilled	Super-heated	Overall
Proposed-original	53.59%	52.38%	81.46%	59.23%
QAC	76.59%	78.41%	18.77%	64.61%
NIQE	76.59%	78.41%	18.77%	64.61%
IL-NIQE	94.67%	50.29%	48.59%	67.24%
SNP-NIQE	95.42%	63.48%	73.69%	78.19%
Q-NIQE	95.79%	68.53%	80.25%	81.74%
CNN-based	95.86%	94.53%	23.65%	79.49%
TL-based	96.13%	94.85%	27.38%	80.53%
Proposed-flame	96.48%	94.16%	90.38%	94.23%

normal condition. In contrast, the proposed method shows consistently good performance (above 90%) under all the conditions. This indicates that the proposed method is more robust and effective than the other methods.

To show the global performance of a model, Fig. 13 illustrates the confusion matrix of the best 4 methods. Some conclusions can be drawn from the figure. First, the normal condition is more easily misclassified as a super-heated condition than the super-chilled condition. The main reason is that the normal and super-heated conditions have higher brightness and clearer edges compared with the super-chilled condition. Second, the super-chilled condition is more easily misclassified as a super-heated condition than the normal condition. This is because the flame videos under normal conditions are of better quality than the flame videos under super-chilled and super-heated conditions.

In many factories, the condition of the kiln is still determined the kilnman by watching the flame video captured by the camera. This is mainly caused by the unstable prediction accuracy. We believe the proposed model will contribute to objective condition recognition. With the objective model, the workload of the kilnman can be significantly reduced. In addition, benefiting from the high prediction accuracy of the proposed sintering recognition model, the energy consumption of the rotary kiln will be greatly reduced. Finally, the proposed model can also be used as a reference for other industrial fields. The proposed idea has great potential in the fractional domain Zhang et al. (2018) and Zhe et al. (2020). For more robust performance, the quality scores of the images/videos may be a good choice for a control system.

5. Conclusion

For more robust visual-based technology applications in industrial systems, we first proposed an application appealing unsupervised BIQA model. Then, this BIQA model is applied to flame images to realize the flame image quality evaluation. Finally, based on the quality score sequence of the flame videos, we proposed a robust condition recognition model. This paper provides the following contributions. In principle, we proposed an unsupervised BIQA method. The proposed model does not need the reference image and subjective quality scores for implementation, and more importantly, has high prediction accuracy and low computational complexity compared with SOTA unsupervised models. Therefore, the proposed model is easy to implement in applications. Regarding application, this paper gives an example of how to use the general BIQA model to finish the industrial flame image quality evaluation. In addition, we provide an example of how to use the quality scores to solve practical industrial problems. In the future, we will further concentrate on the application of IQA in other fields, such as image defogging and rain removal tasks.

CRediT authorship contribution statement

Leyuan Wu: Conceptualization, Writing – original draft, Methodology. **Xiaogang Zhang:** Supervision, Funding acquisition. **Hua Chen:** Project administration, Writing – review & editing. **Yicong Zhou:** Investigation, Software. **Lianhong Wang:** Formal analysis, Resources. **Dingxiang Wang:** Data curation, Validation.

Funding

This paper is supported by the National Key R&D Program of China under Grant 2018YFB1305900 and the National Natural Science Foundation of China under Grant 62171184.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Chan, Kit Yan, Engelke, Ulrich, 2015. Fuzzy regression for perceptual image quality assessment. Eng. Appl. Artif. Intell. 43, 102–110.
Chen, Hua, Zhang, Xiaogang, Hong, Pengyu, Hu, Hongping, Yin, Xiang, 2016. Recognition of the temperature condition of a rotary kiln using dynamic features of a series of blurry flame images. IEEE Trans. Ind. Inf. 12 (1), 148–157.
Fang, Y., Ma, K., Wang, Z., Lin, W., Fang, Z., Zhai, G., 2015. No-reference quality assessment of contrast-distorted images based on natural scene statistics. IEEE Signal Process. Lett. 22 (7), 838–842.
Fong, Cher Min, Wang, Hui Wen, Kuo, Chien Hung, Hsieh, Pei Chun, 2019. Image quality assessment for advertising applications based on neural network. J. Vis. Commun. Image Represent. 63 (Aug.), 102593.1–102593.7.
Gu, Ke, Zhai, Guangtao, Lin, Weisi, Liu, Min, 2016. The analysis of image contrast: From quality assessment to automatic enhancement. IEEE Trans. Cybern. 46 (1), 284–297.
He, K., Sun, J., Tang, X., 2013. Guided image filtering. IEEE Trans. Pattern Anal. Mach. Intell. 35 (6), 1397–1409.
Hua Chen, Xiaogang Zhang, 2020. Burning condition recognition of rotary kiln based on spatiotemporal features of flame video. Energy.
Larson, E.C., Chandler, D., 2010. Categorical image quality (csiq) database. Online, http://vision.okstate.edu/csiq.
Lasmar, N., Stitou, Y., Berthoumieu, Y., 2009. Multiscale skewed heavy tailed model for texture analysis. In: 2009 16th IEEE International Conference on Image Processing. ICIP. pp. 2281–2284.
Le Callet, Patrick, Autrusseau, Florent, 2005. Subjective quality assessment irccyn/IVC database. http://www.irccyn.ec-nantes.fr/ivcdb/.
Li, S.T., Wang, Y.N., Zhang, C.F., 2002. Neural network control system for rotary kiln based on features of combustion flame.

Liu, Y., Gu, K., Zhang, Y., Li, X., Zhai, G., Zhao, D., Gao, W., 2019. Unsupervised blind image quality evaluation via statistical measurements of structure, naturalness and perception. IEEE Trans. Circuits Syst. Video Technol. 1.

Ma, Kede, Liu, Wentao, Zhang, Kai, Duanmu, Zhengfang, Wang, Zhou, Zuo, Wangmeng, 2018. End-to-end blind image quality assessment using deep neural networks. IEEE Trans. Image Process. PP (3), 1.

Mittal, A., Moorthy, A.K., Bovik, A.C., 2012. No-reference image quality assessment in the spatial domain. IEEE Trans. Image Process. 21 (12), 4695–4708.

Narwaria, M., Lin, W., 2012. SVD-based quality metric for image and video using machine learning. IEEE Trans. Syst. Man Cybern. B 42 (2), 347–364.

Ponomarenko, N., Ieremeiev, O., Lukin, V., Egiazarian, K., Jin, L., Astola, J., Vozel, B., Chehdi, K., Carli, M., Battisti, F., Kuo, C.-. J., 2013. Color image database TID2013: Peculiarities and preliminary results. In: European Workshop on Visual Information Processing. EUVIP. pp. 106–111.

Preiss, Jens, Fernandes, Felipe, Urban, Philipp, 2014. Color-image quality assessment: From prediction to optimization. IEEE Trans. Image Process. 23 (3), 1366–1378.

Qiu, Tian, Liu, Minjian, Zhou, Guiping, Wang, Li, Gao, Kai, 2019. An unsupervised classification method for flame image of pulverized coal combustion based on convolutional auto-encoder and hidden Markov model. Energies 12 (13), 2585.

Sharifi, K., Leon-Garcia, A., 1995. Estimation of shape parameter for generalized Gaussian distributions in subband decompositions of video. IEEE Trans. Circuits Syst. Video Technol. 5 (1), 52–56.

Sheikh, L.C.H.R., Wang, Z., Bovik, A.C., 2006. Live image quality assessment database release 2. http://live.ece.utexas.edu/research/Quality/subjective.htm/.

Shnayderman, A., Gusev, A., Eskicioglu, A.M., 2006. An SVD-based grayscale image quality measure for local and global assessment. IEEE Trans. Image Process. 15 (2), 422–429.

Tang, Chongwu, Yang, Xiaokang, Zhai, Guangtao, 2015. Noise estimation of natural images via statistical analysis and noise injection. IEE Trans. Circuits Syst. Video Technol. 25 (8), 1283–1294.

Wang, Zhou, 2004. Image quality assessment : From error visibility to structural similarity. IEEE Trans. Image Process..

Wang, Z., 2011. Applications of objective image quality assessment methods [applications corner]. IEEE Signal Process. Mag. 28 (6), 137–142.

Wang, Zhichao, Liu, Min, Dong, Mingyu, Wu, Lian, 2017. Riemannian alternative matrix completion for image-based flame recognition. IEEE Trans. Circuits Syst. Video Technol. 27 (11), 2490–2503.

Wang, Dingxiang, Zhang, Xiaogang, Chen, Hua, Zhou, Yicong, Cheng, Fanyong, 2020a. A sintering state recognition framework to integrate prior knowledge and hidden information considering class imbalance. IEEE Trans. Ind. Electron. PP (99), 1.

Wang, Dingxiang, Zhang, Xiaogang, Chen, Hua, Zhou, Yicong, Cheng, Fanyong, 2020b. Sintering conditions recognition of rotary kiln based on kernel modification considering class imbalance. ISA Trans. 106, 271–282.

Wu, Q., Wang, Z., Li, H., 2015. A highly efficient method for blind image quality assessment. In: 2015 IEEE International Conference on Image Processing. ICIP. pp. 339–343.

Wu, Leyuan, Zhang, Xiaogang, Chen, Hua, 2019. Effective quality metric for contrast-distorted images based on SVD. Signal Process., Image Commun. 78, 254–262.

Wu, Leyuan, Zhang, Xiaogang, Chen, Hua, Wang, Dingxiang, Deng, Jingfang, 2021. VP-NIQE: An opinion-unaware visual perception natural image quality evaluator. Neurocomputing 463, 17–28.

Wu, Leyuan, Zhang, Xiaogang, Chen, Hua, Zhou, Yicong, 2020. Unsupervised quaternion model for blind colour image quality assessment. Signal Process. 176, 107708.

Xue, W., Zhang, L., Mou, X., 2013. Learning without human scores for blind image quality assessment. In: 2013 IEEE Conference on Computer Vision and Pattern Recognition. pp. 995–1002.

Yousaf, Saqib, Qin, Shiyin, 2015. Closed-loop restoration approach to blurry images based on machine learning and feedback optimization. IEEE Trans. Image Process. 24 (12), 5928–5941.

Zhang, Z., Zhang, J., Ai, Z., 2018. A novel stability criterion of the time-lag fractional-order gene regulatory network system for stability analysis. Commun. Nonlinear Ence Numer. Simul. S1007570418301862.

Zhang, L., Zhang, L., Bovik, A.C., 2015. A feature-enriched completely blind image quality evaluator. IEEE Trans. Image Process. 24 (8), 2579–2591.

Zhang, Teng, Zhou, Zhi-Hua, 2017. In: Precup, Doina, Teh, Yee Whye (Eds.), Multi-Class Optimal Margin Distribution Machine. In: Proceedings of Machine Learning Research, vol. 70, PMLR, International Convention Centre, Sydney, Australia, pp. 4063–4071.

Zhe, Z., Ushio, T., Ai, Z., Jing, Z., 2020. Novel stability condition for delayed fractional-order composite systems based on vector Lyapunov function. Nonlinear Dynam. 99 (2).